

978-9934-564-72-7



# THE LEGAL IMPLICATIONS OF MALICIOUS EXPLOITATION OF SOCIAL MEDIA

Published by the  
NATO Strategic Communications  
Centre of Excellence



ISBN: 978-9934-564-72-7  
Author: Ēriks K. Selga  
Project Manager: Vineta Mēkone  
Copy Editing: Kārlis Streips  
Design: Kārlis Ulmanis

Riga, April 2020  
NATO STRATCOM COE  
11b Kalciema Iela  
Riga LV1048, Latvia  
[www.stratcomcoe.org](http://www.stratcomcoe.org)  
Facebook/stratcomcoe  
Twitter: @stratcomcoe

Eriks K. Selga is a PhD candidate in Law at the University of Hong Kong, where he studies the relationship between regulation and digital technology. He is also an associate scholar at the Foreign Policy Research Institute, a visiting researcher at the Latvian Institute of International Affairs and a student fellow at the Asian Institute of International Financial Law.

This publication does not represent the opinions or policies of NATO or NATO StratCom COE.

© All rights reserved by the NATO StratCom COE. Reports may not be copied, reproduced, distributed or publicly displayed without reference to the NATO StratCom COE. The views expressed here do not represent the views of NATO.

# Table of Contents

Introduction .....	3
--------------------	---

Internet Intermediary Governance Frameworks .....	5
---	---

EU and US Governance Framework	
Trajectories .....	6
The German Experience .....	7
France and Italy .....	11
Shifts in European Praxis on Active Monitoring and Filtering of Content .....	14
Rules on Social Media Intermediaries	
Beyond NATO countries .....	15

Analysis and Recommendations .....	17
------------------------------------	----

Bookmarks .....	20
-----------------	----





*The aim of this contribution is to outline the types of legal frameworks that have been set up by sovereigns to maneuver through and against the malicious use of social media networks, comment on the challenges faced, and identify policy trajectories. Focus is placed on the German Network Enforcement law (NetzDG) 1 as the prototypical archetype for a comprehensive and binding regime for social media intermediaries. Through a transatlantic comparison with other jurisdictions and courts, the legal tendencies of the malicious use of digital space are outlined, and recommendations are provided for the path forward.*

# INTRODUCTION

The growing use of digital media is a well-observed phenomenon, with 90% of adults regularly accessing the Internet, and youth use averaging at least six hours a day.<sup>2</sup> Together with the increasing digitalization of private and public sector models, digital space is creating a parallel space of social activity.<sup>3</sup> This activity functions in a decentralized structure and across a variety of Websites and platforms, each under its own legal regimes and stakeholders.<sup>4</sup> Information flows are filtered through a few 'points of control,' directly impacting interaction between individuals, sovereigns, and other entities. Through popular use, social media and video platforms are becoming especially important gatekeepers. Such platforms have morphed into concentrated arenas for public discourse and attention. While this brings many benefits, it also presents a gamut of new digital manipulation threats which necessitate governance.

For example, the difficulty of authentication has given rise to the use of troll and cyborg entities capable of increasingly authentic proliferation of disinformation narratives. Traditional hacking tools resulting in impersonation capacity are further benefiting from advances in image and sound manipulation software capable of 'deepfaking' individuals. These tools are being amalgamated by states and non-

states to engage in massive social media manipulation and social engineering campaigns.<sup>5</sup> Concurrently, new over- and underground markets have sprawled to collect and broker internet user data, expediting access to information that can be used for the purpose of manipulation.

The initial difficulty with managing the transnational digital domain is only reinforced by the proprietary nature of social media entities. This has made regulating against the malicious use of digital space a complex matter, interweaving several types of stakeholders.

While from a legal standpoint, activity on the Internet is generally not differentiated from activity offline, its digital character necessitates a different approach. The aim of this report is to outline the types of legal frameworks that have been set up by sovereigns to maneuver through and against the malicious use of social media networks, comment on the challenges faced, and identify policy trajectories. Focus is placed on the German Network Enforcement law (*NetzDG*)<sup>6</sup> as the prototypical archetype for a comprehensive and binding regime for social media intermediaries. Through a transatlantic comparison with other jurisdictions and courts, the legal tendencies of the malicious use of digital space are outlined, and recommendations are provided for the path forward.



# Internet Intermediary Governance Frameworks

**The approach to Internet intermediaries differs on each side of the Transatlantic Alliance. In the US, intermediary liability is regulated by two acts – the Communication Decency Act (CDA) of 1996 and the Digital Millennium Copyright Act (DMCA/OCILLA) of 1998.** Together, they provide that Internet intermediaries, including search engines and collaborative platforms, are exempted from liability for the illegal behavior of their users, nor can they be granted injunctions or orders to prevent or terminate illegal user activity. Concurrently, the intermediary is granted immunity from claims for acting against material considered objectionable. According to the DMCA, non-liability to hosts which are infringing is afforded only when there is no knowledge of the infringing material, financial benefit is not received from it, and it is expeditiously removed upon notice.

US courts have generally disallowed direct content-based censorship,<sup>7</sup> and the government has instead used a proxy methodology, wherein a variety of gatekeeping intermediaries are entrusted with the task.<sup>8</sup> These have taken shape in the form of intermediary copyright liabilities, mandatory filters, compelled disclosures of user data, and collateral censorship.<sup>9</sup> Overall, social media in the United States

are likely to enjoy multiple forms of rights to censor speech. They can be protected from compelled speech,<sup>10</sup> even in the form of conditions for grants.<sup>11</sup> The government would also face difficulties in requiring forfeiture of First Amendment rights due to opening the platform to other users, having a dominant market presence, or being considered utilities.<sup>12</sup>

**The EU rules on Internet intermediaries first came into force with the eCommerce Directive (2000/31/EC).** The Directive generally exempted intermediaries that provide conduit, caching and hosting services from secondary liability as long as they are aware that they do not host illegal content or activities.<sup>13</sup> While take-down procedures exist, they are vague and require “expeditious” removal upon “obtaining such knowledge or awareness” of illegal activity.<sup>14</sup> The eCommerce Directive also specifically prohibits the imposition of a general obligation for providers of hosting, caching, and conduit services to monitor the information they transmit or store, and in particular, to “actively seek facts to indicate illegal activity.”<sup>15</sup> A country-of-origin principle is also created to discern the subject of conducted activities under the Directive. Some room is also provided for Member States to establish procedures governing



the prevention and removal of access to such information.

## EU and US Governance Framework Trajectories

These frameworks have set the paradigm for the rules of social intermediaries for nearly the past two decades. They generally have exemption from secondary liability stemming from the illegal activities of their users for reasons of activity promotion, business model preservation, or the prevention of collateral censorship. However, as the weight of social media in steering social discourse has increased, countries have sought to scale up accountability and transparency. Studies have also found that efforts by major social media companies to self-regulate are lacking. For example, the removal of extremist content has decreased over time without the threat of sanction.<sup>16</sup>

Yet any form of binding obligations brings forth myriad questions, with issues of balance of power and burdens between platforms, users, and public authorities. Shifting moderating and monitoring requirements to major social media entities carries the risk of alienating smaller social media for which such requirements are too costly. Automatic filtering systems or live moderators, for instance, may not be a scalable solution for smaller companies that need bespoke approaches because of the language or other characteristics of their customer base. This may place the digital ecosystem in favor of larger

players, lessening pluralism in the market. However, subjecting social media to fewer requirements generally may leave users vulnerable to manipulation.

The EU has championed non-binding efforts to accelerate certain content modification.<sup>17</sup> Its Action Plan and Code of Practice Against Disinformation, for example, was developed to promote self-regulation. It provides more resources and personnel for the European External Action Service and the strengthening of three Strategic Communications Task Forces specifically designed to combat foreign disinformation. It also creates a common Rapid Alert System aimed at facilitating information sharing and coordinating responses to disinformation campaigns. While these measures have been found to curb disinformation, the code lacks enforcement and sanctioning capacity, and it is difficult to quantify the extent of its success.<sup>18</sup>

The system of trickle-down proxy censorship is also challenged by vagueness and uncertainty of the entities' rules, especially given the varying risk appetites towards the type of user engagement that is attracted by the platform.<sup>19</sup> Twitter's prohibition against fear-inciting behavior is different than direct attacks on protected characteristics, though both sets of rules are of thematic equivalence to their objectives. This has led to significant inconsistency in the enforcement of various policies, especially in relation to hate speech and harassment, with the use of fake images and doxing



*Shifting moderating and monitoring requirements to major social media entities carries the risk of alienating smaller social media for which such requirements are too costly. Automatic filtering systems or live moderators, for instance, may not be a scalable solution for smaller companies that need bespoke approaches because of the language or other characteristics of their customer base. This may place the digital ecosystem in favor of larger players, lessening pluralism in the market. However, subjecting social media to fewer requirements generally may leave users vulnerable to manipulation.*



becoming an increasingly difficult aspect of adjudication, in particular when it comes to transparency in terms of explanations and justifications. There is also a trend of using automated tools or setting high sensitivity on the filtering tools that may forego contextually appropriate content – some entities have started requesting users to provide supplemental contextual details.<sup>20</sup>

In line with these worries, US law remains patchwork towards malicious use of digital space. There are isolated islands of protected content, such as certain types of data.<sup>21</sup> However, issues like fake news or impersonation lack a legislative heuristic. Individual cases highlight the trajectory towards forming a praxis at a state level. For example, in 2019, the New York State Attorney General settled a case against a company which was selling fake followers and ‘likes’ to social media account users, because such behavior was deceptive over the use of accounts that impersonate real people, for both computer-operated accounts and human-operated accounts.<sup>22</sup> California, on the other hand, has enacted legislation requiring actors using bots in content online promotions to disclose which posts are machine-generated.<sup>23</sup>

Various discussions are currently taking place on how to proceed further with the regulation of intermediaries and protect against the malicious use of the Internet. There have been proponents for adding limited exceptions to the Section 230 liability shield.<sup>24</sup> Another option would be to change

the scope of immunity towards the EU’s eCommerce Directive. Immunity could also be conditionally removed, if intermediaries do not fulfill the duty of care towards content moderation, in the form of the German NetzDG – currently the most comprehensive binding framework for tackling the malicious abuse of intermediaries.

## The German Experience

The German NetzDG law sets an example of a framework which is targeted only at larger social media. The statute was drafted after requests for voluntary compliance from social media failed,<sup>25</sup> and in fear of disinformation campaigns being able to influence the Bundestag elections in late 2017, it was drafted in a few months.<sup>26</sup> The regulator purposely aimed to target only the top ten largest social media platforms operating in Germany, as it applies to services with at least two million registered users in the country.<sup>27</sup> The law applies to telemedia service providers that:

*“for profit-making purposes, operate Internet platforms which are designed to enable users to share any content with other users to make such content available to the public.”<sup>28</sup>*

This also serves as a broad definition of a social network, which is increased in breadth by the narrow exclusion of two types of services; platforms offering journalistic or editorial content, and platforms aimed

at enabling individual communication or dissemination of content.

NetzDG prohibits unlawful content which breaches the provisions of certain sections of the German Criminal Code – in particular those dealing with offenses against the democratic constitutional state, public order, a person's honor and sexual self-determination. No new definitions or categories of illegal content are created, instead NetzDG differentiates illegal content by degree of detectability. Content which is "manifestly" unlawful, can be detected within 24 hours without "an in-depth examination and with reasonable efforts, i.e., immediately by trained personnel".<sup>29</sup> Content which does not appear immediately unlawful must be removed, generally, within seven days, unless more investigation is necessary. Once the platform receives more than 100 complaints per year, it must publish reports detailing its content moderation praxis on a semi-annual basis.<sup>30</sup> Given that Facebook alone, for example, makes more than 100,000 content-related choices per month among German users alone, the clause exists to further frame the nature of a 'social network'.

The framework also implements a complaint procedure that should be easily and directly accessible to all users, and the mechanism must ensure that the social network provider assesses the complaint "immediate[ly]".<sup>31</sup> Decisions on unlawfulness must be referred to a "recognized self-regulation institution" where:

- Independence of the expertise of analysts is ensured;
- Facilities are in place to appropriately analyze the complaint within seven days;
- Rules of procedure regulate the criteria and scope of submission and decision requirements, allowing for decisions to be reviewed;
- A complaints service mechanism has been created;
- The institution is funded by multiple social network providers and is open to admitting new providers.

The German Telemedia Act of 2007 has also been amended to allow for the disclosure of subscriber data for the enforcement of civil law claims arising from the violation of protected rights under NetzDG – this disclosure requires a court order. If a social media network negligently violates the rules, obligations can be set, with fines ranging from up to EUR 5 million for failing to react to a complaint, to up to EUR 50 million for all other violations. The Ministry of Justice, however, must first obtain a court decision declaring the relevant content to be illegal.<sup>32</sup>

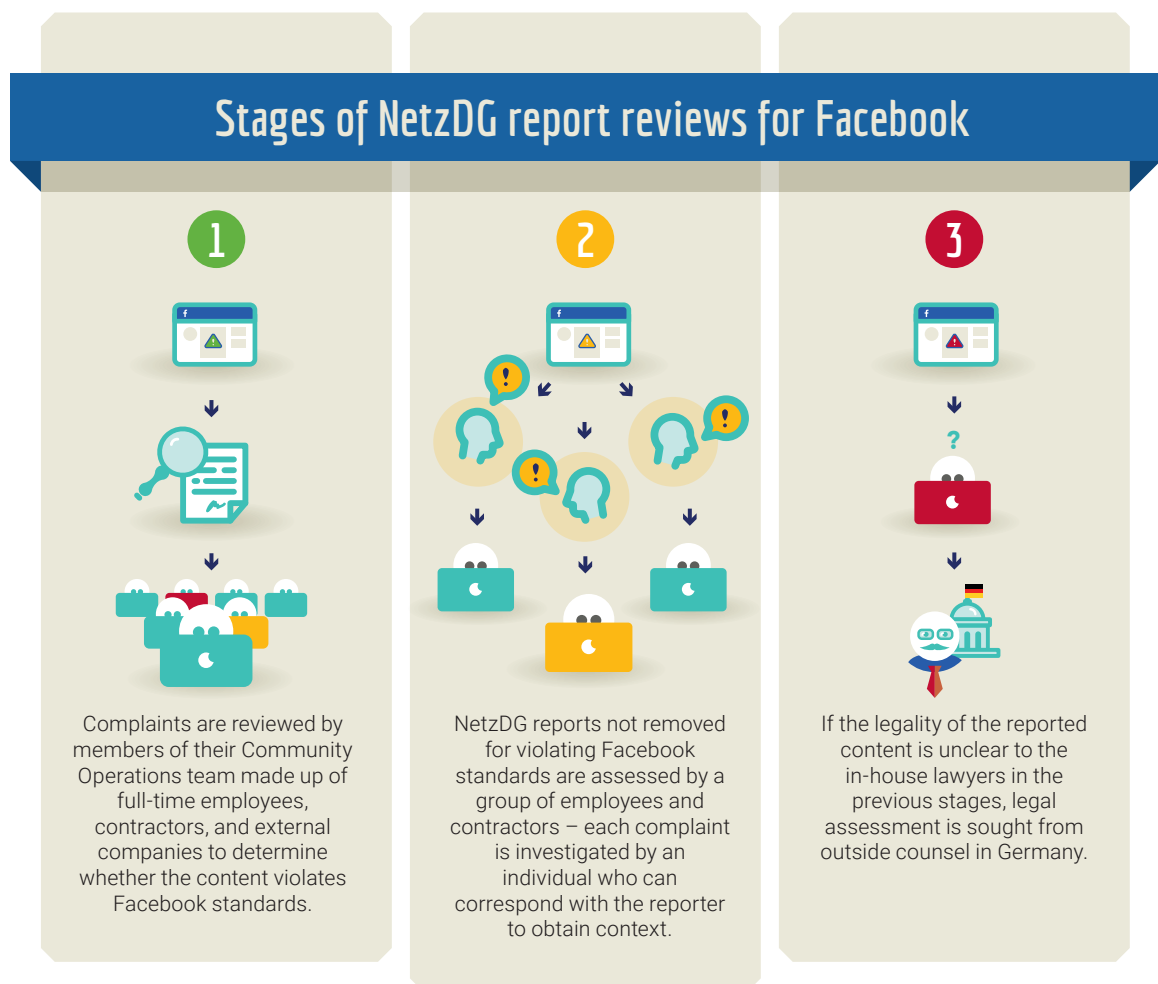
When it comes to Facebook, the NetzDG reports are reviewed in three stages.<sup>33</sup> First, complaints are reviewed by members of their Community Operations team made



up of full-time employees, contractors, and external companies to determine whether the content violates Facebook standards. Second, NetzDG reports not removed for violating Facebook standards are assessed by a group of employees and contractors – each complaint is investigated by an individual who can correspond with the reporter to obtain context. Third, if the legality of the reported content is unclear to the in-house lawyers in the previous stages,

legal assessment is sought from outside counsel in Germany. The reporting party and reported user are then informed about the decision.

For Google and Facebook, the independent institution they work with is the Voluntary Self-Control for Multimedia Service Providers association, a non-profit that has been working with protecting the rights of young people in online media since 2003,



by ensuring that online media content is not harmful to their development or illegal.<sup>34</sup> They help Google and Facebook arbitrate a handful of cases every month that require external counsel.<sup>35</sup>

Importantly, NetzDG does not address re-uploads of content – content must be flagged anew even if it is identical to previous uploads.<sup>36</sup> This change was inserted to ensure compliance with the eCommerce Directive’s prohibition on proactive content search requirements.


The first years of the law have been contentious. German voters polled gave a strong approval rating of 87%, with 5% disapproving. A multitude of other interest groups have criticized the law for the state abdicating its responsibility and giving social media platforms a judicial role in deciding the illegality of a role.<sup>37</sup> Advocates of free expression posited concern regarding the indirect pressure on social networks to over-comply.<sup>38</sup> In a 2018 report, the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression posited that the government should avoid heavy-handed regulation with disproportionate sanctions on intermediaries and avoid proactive filtering of content, while avoiding the delegation of authority to corporate entities or government agencies instead of judicial authorities. The Special Rapporteur found that NetzDG prohibition against the dissemination of information on criteria like defamation does not demand the level

of protection afforded to them, and thus is incompatible with Article 19 of the ICCPR.<sup>39</sup>

The results of this law are unclear. While many pieces of content have been removed, their removal was overwhelmingly based on the proprietary community standards of the social network, which existed before. Lawsuits against Facebook for allegedly incorrectly deleted content have challenged the actions in terms of the content being compliant with German law, interpreting hate-speech too broadly, or over-interpreting the freedom of expression too narrowly.<sup>40</sup> Given the very context-specific nature of each case, the rulings have not manifested in a unified stance by courts, and in some cases have even been contradictory.<sup>41</sup> Facebook was fined EUR 2 million earlier in 2019 for failing to be adequately transparent about the complaints received about illegal material, but it is appealing the decision.<sup>42</sup> The German government aims to publish a study of NetzDG’s impact in a few years.<sup>43</sup>

Germany has also spearheaded efforts to pluralize social media by proposing a bill aimed at binding their ranking and sorting algorithms. The bill is a result of significant discourse about the dominance of social media platforms not just in the market, but also in their own forums, and about the fact that their business interests may not lead to the most optimally democratic outcomes.<sup>44</sup> The bill obliges video platforms such as Netflix and intermediaries like social media platforms and search engines to disengage unfair hindering or promotion of content





*The results of this law are unclear. While many pieces of content have been removed, their removal was overwhelmingly based on the proprietary community standards of the social network, which existed before. Lawsuits against Facebook for allegedly incorrectly deleted content have challenged the actions in terms of the content being compliant with German law, interpreting hate-speech too broadly, or over-interpreting the freedom of expression too narrowly.*

in terms of access, search, and browsing features for users. They are mandated to prioritize public broadcasting content and offer at least two sorting modes (such as alphabetical or chronological). While media intermediaries do not have the same search obligations, they have to ensure that providers of journalistic editorial content are not discriminated to the point of having a significant influence on its visibility.<sup>45</sup>

Both types of platforms are also responsible for disclosing in a simple, recognizable and directly accessible format the selection criteria for sorting and presenting content. The function of the algorithm used to search and present content must be disclosed for the weight it provides for different criteria, also explaining reasons that motivate content suggestions.<sup>46</sup>

## France and Italy

The German law is comprehensive, but it purposely avoids creating new categories of acts or content, instead opting for the existing classes in its criminal law. Other governments, heavily drawing from the German experience, have modified their laws to include such definitions.

The French law against the manipulation of information, adopted in July 2019, defines information manipulation as:

*"inexact or misleading allegation of a fact that could alter the sincerity of an upcoming vote*

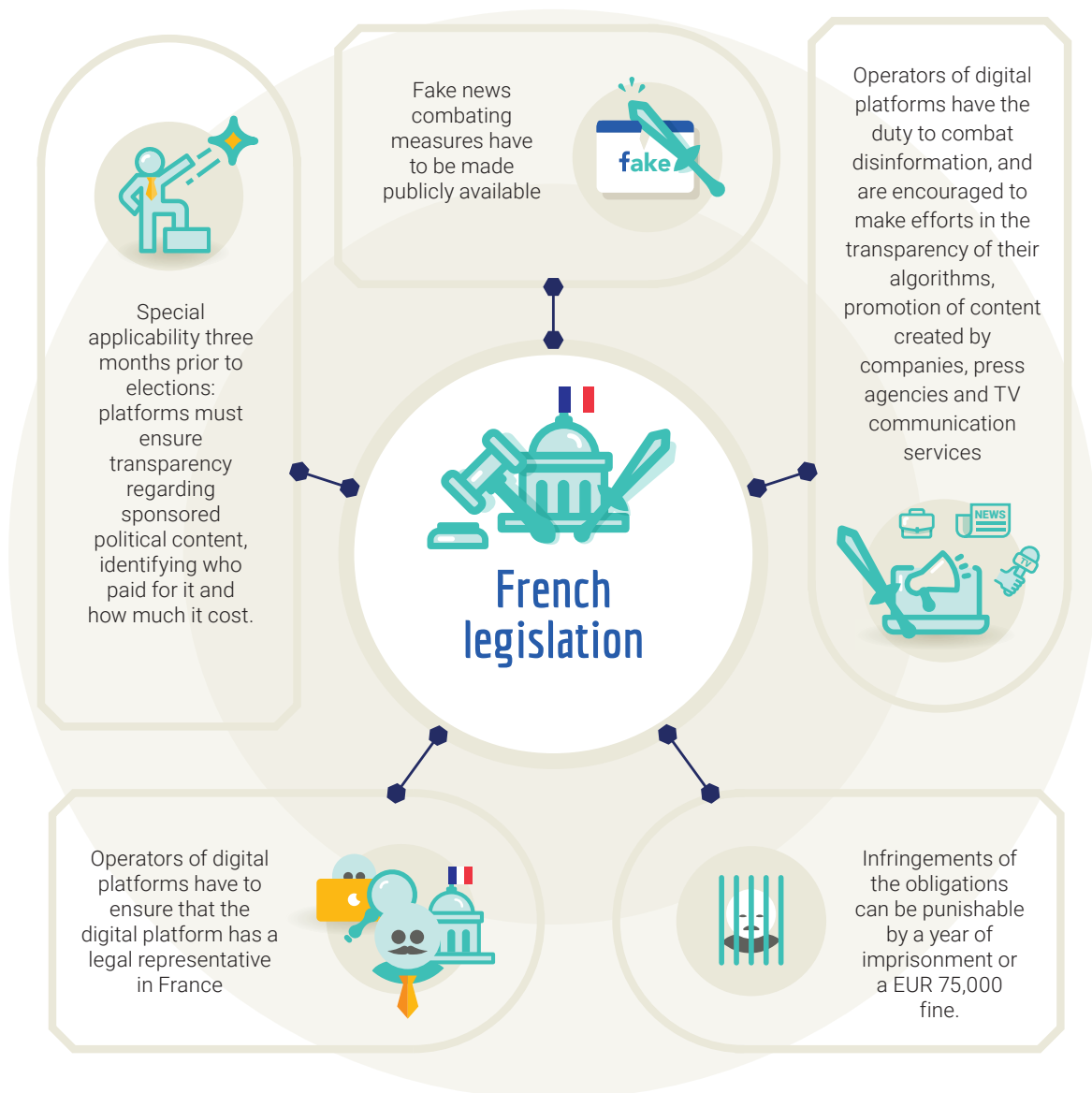
*and that is spread deliberately, artificially or automatically and massively to the online public through a communication service"*<sup>47</sup>

The law states that operators of digital platforms have the duty to combat disinformation, and that they are encouraged to make efforts in the transparency of their algorithms, promotion of content created by companies, press agencies and TV communication services, the fight against accounts diffusing false information, and among others, to ensure that the digital platform has a legal representative in France. The law also has special applicability three months prior to elections. During this time period, platforms must ensure transparency regarding sponsored political content, identifying who paid for it and how much it cost.<sup>48</sup> Any individual, political party, association or candidate can petition a judge to call for the removal of the content. The judge has 48 hours to decide whether the content should be removed if it meets the following criteria:

1. The news being manifest
2. It is being disseminated deliberately on a large scale
3. It has led to a disturbance of peace or has compromised election results.

Content platforms must introduce fake news-combating measures and make them publicly available. Infringements of the





obligations can be punishable by a year of imprisonment or a EUR 75,000 fine.

In a separate bill focused on online hate speech, France introduced requirements for platforms to create reporting systems which allow users to submit notification of illegal content which is related to race, gender, religion, sexual orientation or disability.<sup>49</sup> The

content must be removed within 24 hours based on a decision by the platform. In case of non-compliance, platforms could face fines of up to 4% of their total revenues, and officials can personally face a year of imprisonment and fines of up to a quarter-million euro. Individuals abusing the reporting mechanisms can also face prison time and monetary fines. The Audiovisual

Council is the authority responsible for overseeing the level of compliance with the mandated rules. However, the European Commission has officially requested that France postpone the adoption of the law, in fear of breaching the requirements of the eCommerce Directive.<sup>50</sup>

Italy, for example, created a website in 2018 by which readers can report suspected disinformation to the Italian State Police, which will then investigate the matter and take legal steps if evidence supports a conclusion that the content is fake news.<sup>51</sup> Depending on whether it is a wide-spread media release, the individual could face up to five years in prison.<sup>52</sup> The bill following the creation of the Website defined fake news as information that is manifestly unfounded or untrue, is clearly harmful to the public interest, or alarms the public. The fake-news law has not advanced out of its parliamentary committee, instead advancing initiatives related to guidelines for social platforms and newspapers.<sup>53</sup>

On the basis of the UN International Covenant on Civil and Political Rights, the Special Rapporteur on Human Rights has noted concern over the protection of freedom of expression in light of opaque investigations conducted by the Italian police.<sup>54</sup> The rapporteur is concerned that the ICCPR Article 19(3) restriction on the right to freedom is not proportional, because as long as the right is provided by law and is necessary for the rights, reputations, safety, security, or health and

morals of others, online restrictions can be the same as off-line restrictions, but they must be sufficiently clear, accessible and predictable. Indeed, any such restrictions must be assessed under the principle of proportionality to ensure that they target a specific objective and do not unduly intrude upon the rights of targeted persons.

The chosen restrictions must also be the least intrusive form to achieve the desired result, which is particularly important in securing public debate in a democratic society. In this regard, the OSCE Joint Declaration on Freedom of Expression and “Fake News” found that general prohibitions on disseminating information based on vague and ambiguous ideas, including “false news” and non-objective information, are “incompatible and should be abolished.”<sup>55</sup>

### Shifts in European Praxis on Active Monitoring and Filtering of Content

A 2014 decision of the European Court of Justice required Google to remove links associated with a person’s name appearing in searches.<sup>56</sup> In 2017 the same court also found that even the indexation of the metadata of protected works via a search function which allows the sharing of those works by other users via peer-to-peer networks engages in communication to the public of those works, and is thus subject to respective sanctions.<sup>57</sup> The European Union has generally recommended that member states not create legal rules for actively monitoring and filtering illegal content.<sup>58</sup>





However, the Court of Justice of the European Union has recently changed its handling of the issue. In October of 2019, when presented with the question of whether Facebook should proactively seek to block defamatory posts that are identical or equivalent in content to materials defined in Article 15 of the eCommerce Directive. The Court found that injunctions covering identical and equivalent content to that which is found illegal by a court are allowed by the Directive, but they must not require the platform to independently examine whether the content violates the law. In this, the Court diverges from the standards set in previous cases to advance the idea that Facebook may be required to monitor every post by every customer, even if it was not specifically specified in advance by a court.<sup>59</sup>

The legal drafts of the Terrorist Content Regulation, however, observe that derivative mandatory proactive measures should not lead to imposing a general monitoring obligation, but derogations may occur.<sup>60</sup> The bill proposes subjecting entities beyond social media networks. Hosting service providers are defined as:

“provider[s] of information society services consisting of the storage of information provided by and at the request of the content provider and in making the information stored available to the public”

Hosting service providers are obliged to take “appropriate, reasonable and proportionate

actions” against the dissemination of terrorist content, with due regard to fundamental rights and the freedom of expression. The competent authority will have the power to issue a decision requiring the hosting service provider to remove the terrorist content within one hour of receipt. They are also responsible for taking proactive measures, including the use of automated tools to prevent the re-upload of content that has previously been flagged or to detect and identify terrorist content. Human oversight and complaint mechanisms are also necessary.<sup>61</sup>

### Rules on Social Media Intermediaries Beyond NATO countries

The use by certain nations of the NetzDG modalities has raised concerns about the ultimate benefactor and purpose of the laws. The 2018 Egyptian Media and Press Law provides the national regulator with the right to suspend any personal Websites, blogs, or social media accounts with more than 5,000 followers if they post “fake news, promote violence, or spread hateful views.”<sup>62</sup> Individuals can be prosecuted for even encouraging others to break the law via their content.

In 2019 Singapore set in force the Protection from Online Falsehoods and Manipulation Act, allowing heads of Singaporean ministries to issue fines and imprison individuals over “falsehoods” against public interests.<sup>63</sup> Notably, the rules do not apply to satire, parody, opinions, or criticisms of

the government and its policies, instead explicitly adhering to restricting false statements of fact.<sup>64</sup> They also grant a range of corrective orders which allow online intermediaries not just to remove, but also to edit content if so requested by the government. This applies not only to social networks, but to any form of mobile networks. Sri Lanka is similarly considering the adoption of amendments to the penal and criminal codes to criminalize the dissemination of “false news” that affects “communal harmony” or state security.<sup>65</sup>

Russia drafted legislation in 2019 to target “unreliable information,” defined as “unreliable, socially significant information disseminated under the guise of reliable messages which creates a threat to life and/or the health of citizens or property, the threat of mass disturbance of public order and/or public safety, or the threat of creating or impairing the proper operation of vital elements of transport or social infrastructure, credit institutions, energy, facilities, industry or communications.”<sup>66</sup> The state authority can act on its own or on the basis of a complaint by a third party to oblige a host to delete the relevant content within 24 hours. The legislation also targets content that expresses “lack of respect for the authorities” and “offends human dignity and public morality, and displays obvious disrespect for society, the state, [and] the official state symbols...”

The People’s Republic of China presents a different type of comprehensive Internet

governance framework. Ever since the idea of ‘Internet Sovereignty’ was raised in China in 2010, the country has sought to develop government capacity to supervise, regulate, and censor all content on the Internet within China’s national borders without interference.<sup>67</sup> The Chinese cybersecurity law which came into effect in 2017 concretized the vision. It applies to all ‘network operators’<sup>68</sup> which own, manage, or provide network services, as well as to virtually any device capable of interfacing with data. People are obliged to obey social norms, accept supervision from the government and the public, and bear social responsibility.<sup>69</sup> Persons or organizations are explicitly prohibited from engaging in a wide range of activities against public order or security, including activities endangering national honor, interests, the socialist system, or national unity. The law also expressly prohibits the proliferation of false information to disrupt economic or social order, or information infringing upon the rights of others.<sup>70</sup>

The law has been implemented extensively. For example, in the months leading up to the 19<sup>th</sup> Communist Party Congress in 2017, foreign media were removed from China’s social media and video platforms, and certain TV shows were labelled as illegal. More broadly, several thousand Websites have been shut down since, and more internal censors have been hired. The People’s Republic’s leading social media platforms regularly shutter accounts considered to be disseminating gossip.



The control is only bound to increase in intensiveness given that China's Cyberspace Administration, the Internet content authority, released a new set of rules a few months ago, which expands on the bans of illegal content.<sup>71</sup> Network operators are, for instance, ordered to restrict content such as exaggerated headlines which could encourage minors to pursue unsafe acts. Recommendations are issued to institute algorithms to promote 'proper' content instead.<sup>72</sup>



# Analysis and Recommendations

Legal initiatives in the West cover a range of issues ranging from fake news to illegal content, and on to information manipulation. The majority of new initiatives are focused on the traditional understanding of social media – Facebook, Twitter, Google – in light of their observed ability to enable disinformation<sup>73</sup> via fake profiles and groups, online advertising and clickbait, and micro-targeting and manipulative use of third-party data analysis. However, the novel legal frameworks do not proceed far beyond the modalities of moderating specific pieces of content. The discussion remains framed around how users, social networks and governments can work together to achieve a fair and democratic outcome regarding singular units of information. Creating procedural frameworks for handling individual cases with high levels of certainty and legitimacy has been a logical priority. Such an approach leaves many gaps for disinformation-fostering behavior to proceed, but it is an important first step toward covering at least the ends of the disinformation campaigns before tackling the means.

The means involve a wide range of tools, ranging from the purposeful dissemination of inciteful opinions, to the abuse of social media algorithms to advance a narrative, to even the malignant and fraudulent use of technological tools for hacking or

impersonation. The vast range of such tools and their different degrees of legality at this point prevent a comprehensive framework. Purchasing followers or likes to support certain content is currently only regulated by the Terms of Service of social media platforms. Ensuring that parties cannot use such a tool without the voluntary participation of platforms would also require cross-checking jurisdictions to the origin of the service provider. The test for legality would have to be made against the relevant system's framework. Troll farms are an even more evasive tool, as attempts to shut down their dissemination of certain narratives would face a direct challenge to the freedom of expression. It is also conversely difficult to account for the innate partialities of individuals who make them the targets or amplifiers of such campaigns, or the motivations for producing and distributing them. Studies often find that online consumption of information and its further proliferation are more of an indicator of membership to certain communities than a search for or even belief in its objectivity.<sup>74</sup>

At the moment, a comprehensive legal framework is unlikely. Important steps have been taken to create a capacity to affect particular content, but major European nations have taken divergent approaches to legislating online behavior, focusing efforts instead on different platforms, sources, and





*At the moment, a comprehensive legal framework is unlikely. Important steps have been taken to create a capacity to affect particular content, but major European nations have taken divergent approaches to legislating online behavior, focusing efforts instead on different platforms, sources, and chronologies.*

chronologies. The US still has not chosen a path for updating its online regime. Further legislative progress is highly likely to vary across jurisdictions, and legal trajectories will concretize only at a national level. This is similar to the multiple rounds of legislative development in other nascent legal regimes like money laundering or virtual currency regimes, where many initiatives are based on trial and error. The iterative legislative process will be critical for creating definitions, praxis, and data for further examination.

While investigation into halting these means of disinformation continues, the law has multiple access points to expand in, so as better to halt the spread of illegal content, including the strengthening of existing legislation on misleading advertising, election silence periods, political spending, consumer rights, and data protection rules.<sup>75</sup> The examined legal documents highlight that the updating of existing norms and bringing them directly into the online flora via connector legislation is a path which faces little resistance. Many options for regulating automated content recognition technology in the area of disinformation have been proposed, ranging from allowing for the continuation of the status quo, to forms of self-regulation with differing extents of audits, co-regulation between governments and industry, and statutory regulations ordering a regulator to combat disinformation directly by licensing or other moderation mechanisms. In the European Union

alone, a litany of policy proposals and discussions have deliberated on the use of new technology in such initiatives, ranging from voluntary compliance programs, codes, principles, and varying degrees of recommendations for algorithmic and AI based content moderation, particularly calling for transparency in their use. The creation of an internet Ombudsman has been proposed at the Council of Europe to assess whether content is legal or illegal, and it could accept questions from Internet intermediaries.<sup>76</sup>

However, critical to the development of legal rules in the online environment is an understanding of the purpose of such goals. The current regime was developed in reaction to fear of interferences in political campaigns, terrorist threats, and other high-level incidents. The resulting mechanisms function as quick-response firemen teams capable of putting out content-based incidents. The approach has been conservative, and various stakeholders have urged for continued restraint to ensure that the legal instruments cannot be used to stifle fundamental freedoms. Governments espousing different values have shown how the mechanisms can be abused, and the worries are warranted.

Countries should continue conducting gap reviews to better understand which ones the law cannot fulfil beyond the aforementioned need to have an emergency instrument. In the meantime, there are three important steps that states should take:



**First, states should confidently increase transparency requirements for Internet entities to amass the data necessary to understand these gaps and their origin.**

This will allow for more tangible analysis of the impact brought by malevolent activity online, which will necessarily be context-specific. States can formulate their priorities and engage with the relevant stakeholders on the basis of these findings.

**Second, states should expand their focus on bringing traditional non-discrimination rules to the digital domain,** so as to develop a more plural digital environment, and ultimately assert which information typologies are purposefully malicious, thus requiring counter activity, or are authentic emanations of people's opinions. This information can further be used across states to develop more comprehensive approaches.

**Third, countries should also begin capacity-building for the digital environment,** as more pertinent gaps in capabilities between states may become a factor for exploitation. An advanced legal framework will be ineffective without appropriate resources for its enforcement. The weakest links in the Transatlantic space can become hotbeds for disinformation propagation, or even data off-shores, impeding the defensive capacity of all states.

While in the short-term, legal drafting and enforcement of malicious internet use will

likely remain a domestic responsibility, international standards should be calibrated concurrently. The various experiences and dimensions nations have taken to counter the malicious use of the Internet provide ample learning opportunities. Standards should be centered on collective, democratic and liberal principles, juxtaposing the emanations of some regimes beyond the Transatlantic. These foundations will ease cooperation and normalize the still early fight against Internet exploiters. Once a unified trajectory and groundwork are secured, a more comprehensive multilateral legal framework can be developed.



# *Next steps:*

- 1. Democratic, liberal and collective principles should be used as legislation standards and as such should be prioritised from the day one*
- 2. Increased transparency requirements for Internet entities to amass the data necessary to understand existing legislation gaps and further legislations needs*
- 3. Governments should undertake their knowledge and capacity-building for the digital environment*
- 4. It will require iterative legislative processes to achieve comprehensive legal framework*



# Endnotes

- 1 Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act), July 2017. See communication No. OL DEU 1/2017.
- 2 Jean M. Twenge, Gabrielle N. Martin, and Brian H. Spitzberg, "Trends in U.S. Adolescents' Media Use, 1976–2016: The Rise of Digital Media, the Decline of TV, and the (near) Demise of Print," *Psychology of Popular Media Culture* 8, no. 4 (2019): 329–45, <https://doi.org/10.1037/ppm0000203>; Jacob Poushter, Caldwell Bishop, and Hanyu Chwe, "Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones," *Pew Research Center Pew Research Center* (n.d.): 46.
- 3 Britt Christensen, "Cyber State Capacity: A Model of Authoritarian Durability, ICTs, and Emerging Media," *Government Information Quarterly* 36, no. 3 (July 1, 2019): 460–68, <https://doi.org/10.1016/j.giq.2019.04.004>; Christine Legner et al., "Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community," *Business & Information Systems Engineering* 59, no. 4 (August 1, 2017): 301–8, <https://doi.org/10.1007/s12599-017-0484-2>; Daniel R. A. Schallmo and Christopher A. Williams, "Digital Transformation of Business Models," in *Digital Transformation Now!: Guiding the Successful Digitalization of Your Business Model*, ed. Daniel R. A. Schallmo and Christopher A. Williams, SpringerBriefs in Business (Cham: Springer International Publishing, 2018), 9–13, [https://doi.org/10.1007/978-3-319-72844-5\\_3](https://doi.org/10.1007/978-3-319-72844-5_3).
- 4 See Tarleton Gillespie, "Regulation of and by platforms," in Jean Burgess, Thomas Poell, and Alice Marwick, eds., *The SAGE Handbook of Social Media* (SAGE Publications, 2017).
- 5 "Computational Propaganda Worldwide: Executive Summary," The Computational Propaganda Project, accessed December 19, 2019.
- 6 Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act), July 2017. See communication No. OL DEU 1/2017.
- 7 See for example *Ashcroft v. ACLU*, 542 U.S. 656, 666 (2004).
- 8 Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 Harv. L. Rev. 2296, 2309 (2014).
- 9 A recent US case held that a private operator of a public access TV channel is not a state actor bound by the First Amendment, even if that government deems it as such. A broad reading of the decision may pivot state action criteria regarding the creation of a public forum towards no longer being dependent on control, but on property-interest in the entity, amounting to a possible violation of a media platform's First Amendment rights. See *Manhattan Community Access Corp. v. Halleck* 139 S. Ct. 1921 (2019).
- 10 *Wooley v. Maynard* and *Board of Education v. Barnette*
- 11 *Agency for Int'l Dev. v. Alliance for Open Soc'y Int'l*
- 12 See *Manhattan Community* and *Miami Herald Pub. Co. v. Tornillo*, and *Pacific Gas and Elec. Co. v. Public Utilities Com'n of California* respectively.
- 13 Electronic Commerce Directive, Directive 200/31/EC of the European Parliament and of the Council
- 14 *Ibid*, Article 14
- 15 *Ibid*. Article 15
- 16 William Echikson and Olivia Knodt, "Germany's NetzDG: A Key Test for Combatting Online Hate," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 22, 2018), <https://papers.ssrn.com/abstract=3300636>. "plainCitation": "William Echikson and Olivia Knodt, "Germany's NetzDG: A Key Test for Combatting Online Hate," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 22, 2018
- 17 Sartor Giovanni, "Providers Liability: From the ECommerce Directive to the Future," In-Depth Analysis for the IMCO Committee (European Parliament, 2017).
- 18 "Tackling Disinformation: Going Beyond Content Moderation," Institut Montaigne, accessed December 18, 2019.
- 19 Sarah Roberts, Content Moderation (University of California at Los Angeles, 2017). See also ARTICLE 19 submission, p. 2
- 20 YouTube policies (the importance of context); Facebook community standards (hate speech)
- 21 *Ibid*.
- 22 "Attorney General James Announces Groundbreaking Settlement With Sellers Of Fake Followers And 'Likes' On Social Media | New York State Attorney General," accessed December 18, 2019.
- 23 See *Cal. Bus. & Prof. Code* § 17940, *et seq.*
- 24 Presentation by Stephen Philip Mulligan, Legislative Attorney, Congressional Research Service, at "Towards Rule of Law in the Digital Environment" Discussion held in Brussels, 11 December 2019
- 25 "Facebook Must Ban Abusive Content, Says German Justice Minister Maas | DW | 27.08.2015," DW.COM, accessed December 17, 2019.
- 26 Wolfgang Schulz, "Regulating Intermediaries to Protect Privacy Online – The Case of the German NetzDG," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, July 19, 2018), <https://papers.ssrn.com/abstract=3216572>. \u0000\u0000\u0000 SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, July 19, 2018



- 27 Heidi Tworek and Paddy Leerssen, "An Analysis of Germany's NetzDG Law," *Working Paper*, Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, n.d., 11.
- 28 "Gesetz Zur Verbesserung Der Rechtsdurchsetzung in Sozialen Netzwerken [Netzwerksdurchsetzungsgesetz] [NetzDG] (Network Enforcement Act or NetzDG)," DEUTSCHER BUNDESRAT: DRUCKSACHEN [BR-Drs.] 536/17 § (2017), Deutscher Bundesrat website.
- 29 Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [Netzwerksdurchsetzungsgesetz] [NetzDG] (Network Enforcement Act or NetzDG).
- 30 Chris Köver, "Facebook verrät, wie viele Hasskommentare es wirklich löscht," *Wired.de*, September 26, 2016.
- 31 Tworek and Leerssen, "An Analysis of Germany's NetzDG Law"; Sabrina Galli, "NYDFS Cybersecurity Regulations: A Blueprint for Uniform State Statute Notes & Comments: Section III: Cybersecurity & Fintech," *North Carolina Banking Institute*, 2018, 3.
- 32 Jenny Gesley, "Germany: Social Media Platforms to Be Held Accountable for Hosted Content Under 'Facebook Act,'" web page, *Global Legal Monitor*, July 11, 2017.
- 33 "NetzDG Transparency Report", Facebook, July 2019.
- 34 "FSM | About Us," accessed December 18, 2019, <https://www.fsm.de/en/about-us>.
- 35 "FSM | About Us."
- 36 Echikson and Knodt, "Germany's NetzDG." or NetzDG law represents a key test for combatting hate speech on the internet. Under the law, which came into effect on January 1, 2018, online platforms face fines of up to €50 million for systemic failure to delete illegal content. Supporters see the legislation as a necessary and efficient response to the threat of online hatred and extremism. Critics view it as an attempt to privatise a new 'draconian' censorship regime, forcing social media platforms to respond to this new painful liability with unnecessary takedowns. This study shows that the reality is in between these extremes. NetzDG has not provoked mass requests for takedowns. Nor has it forced internet platforms to adopt a 'take down, ask later' approach. At the same time, it remains uncertain whether NetzDG has achieved significant results in reaching its stated goal of preventing hate speech. This paper begins by explaining the background that led to the development and passage of NetzDG. It examines the reaction to the law by civil society, platforms and the government. It concludes with suggestions, for platforms, civil society and the authorities, on ways to improve the law to be effective in the fight against online hate while keeping the internet open and free. CEPS acknowledges the Counter Extremism Project's support for this research. The study was conducted in complete independence. It is based on interviews with regulators, company representatives, and civil society activists. The authors take full responsibility for its findings.", "event-place": "Rochester, NY", "genre": "SSRN Scholarly Paper", "language": "en", "number": "ID 3300636", "publisher": "Social Science Research Network", "publisher-place": "Rochester, NY", "source": "papers.ssrn.com", "title": "Germany's NetzDG: A Key Test for Combatting Online Hate", "title-short": "Germany's NetzDG", "URL": "https://papers.ssrn.com/abstract=3300636", "author": [{"family": "Echikson", "given": "William"}, {"family": "Knodt", "given": "Olivia"}], "accessed": {"date-parts": [{"2019", 12, 17}], "issued": {"date-parts": [{"2018", 11, 22]}}}, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}- 37 Echikson and Knodt. or NetzDG law represents a key test for combatting hate speech on the internet. Under the law, which came into effect on January 1, 2018, online platforms face fines of up to €50 million for systemic failure to delete illegal content. Supporters see the legislation as a necessary and efficient response to the threat of online hatred and extremism. Critics view it as an attempt to privatise a new 'draconian' censorship regime, forcing social media platforms to respond to this new painful liability with unnecessary takedowns. This study shows that the reality is in between these extremes. NetzDG has not provoked mass requests for takedowns. Nor has it forced internet platforms to adopt a 'take down, ask later' approach. At the same time, it remains uncertain whether NetzDG has achieved significant results in reaching its stated goal of preventing hate speech. This paper begins by explaining the background that led to the development and passage of NetzDG. It examines the reaction to the law by civil society, platforms and the government. It concludes with suggestions, for platforms, civil society and the authorities, on ways to improve the law to be effective in the fight against online hate while keeping the internet open and free. CEPS acknowledges the Counter Extremism Project's support for this research. The study was conducted in complete independence. It is based on interviews with regulators, company representatives, and civil society activists. The authors take full responsibility for its findings.", "event-place": "Rochester, NY", "genre": "SSRN Scholarly Paper", "language": "en", "number": "ID 3300636", "publisher": "Social Science Research Network", "publisher-place": "Rochester, NY", "source": "papers.ssrn.com", "title": "Germany's NetzDG: A Key Test for Combatting Online Hate", "title-short": "Germany's NetzDG", "URL": "https://papers.ssrn.com/abstract=3300636", "author": [{"family": "Echikson", "given": "William"}, {"family": "Knodt", "given": "Olivia"}], "accessed": {"date-parts": [{"2019", 12, 17}], "issued": {"date-parts": [{"2018", 11, 22]}}}, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json"



<https://github.com/citation-style-language/schema/raw/master/csl-citation.json>

- 38 Echikson and Knodt. or NetzDG law represents a key test for combatting hate speech on the internet. Under the law, which came into effect on January 1, 2018, online platforms face fines of up to €50 million for systemic failure to delete illegal content. Supporters see the legislation as a necessary and efficient response to the threat of online hatred and extremism. Critics view it as an attempt to privatise a new 'draconian' censorship regime, forcing social media platforms to respond to this new painful liability with unnecessary takedowns. This study shows that the reality is in between these extremes. NetzDG has not provoked mass requests for takedowns. Nor has it forced internet platforms to adopt a 'take down, ask later' approach. At the same time, it remains uncertain whether NetzDG has achieved significant results in reaching its stated goal of preventing hate speech. This paper begins by explaining the background that led to the development and passage of NetzDG. It examines the reaction to the law by civil society, platforms and the government. It concludes with suggestions, for platforms, civil society and the authorities, on ways to improve the law to be effective in the fight against online hate while keeping the internet open and free. CEPS acknowledges the Counter Extremism Project's support for this research. The study was conducted in complete independence. It is based on interviews with regulators, company representatives, and civil society activists. The authors take full responsibility for its findings. "event-place": "Rochester, NY", "genre": "SSRN Scholarly Paper", "language": "en", "number": "ID 3300636", "publisher": "Social Science Research Network", "publisher-place": "Rochester, NY", "source": "papers.ssrn.com", "title": "Germany's NetzDG: A Key Test for Combatting Online Hate", "title-short": "Germany's NetzDG", "URL": "https://papers.ssrn.com/abstract=3300636", "author": [{"family": "Echikson", "given": "William"}, {"family": "Knodt", "given": "Olivia"}], "accessed": {"date-parts": [{"2019, 12, 17}], "issued": {"date-parts": [{"2018, 11, 22}]}}], "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json"
- 39 David Kaye, "Letter of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression to the German Government," June 1, 2017.
- 40 Echikson and Knodt, "Germany's NetzDG." or NetzDG law represents a key test for combatting hate speech on the internet. Under the law, which came into effect on January 1, 2018, online platforms face fines of up to €50 million for systemic failure to delete illegal content. Supporters see the legislation as a necessary and efficient response to the threat of online hatred and extremism. Critics view it as an attempt to privatise a new 'draconian' censorship regime, forcing social media platforms to respond to this new painful liability with unnecessary takedowns. This study shows that the reality is in between these extremes. NetzDG has not provoked mass requests for takedowns. Nor has it forced internet platforms to adopt a 'take down, ask later' approach. At the same time, it remains uncertain whether NetzDG has achieved significant results in reaching its stated goal of preventing hate speech. This paper begins by explaining the background that led to the development and passage of NetzDG. It examines the reaction to the
- 41 Echikson and Knodt. or NetzDG law represents a key test for combatting hate speech on the internet. Under the law, which came into effect on January 1, 2018, online platforms face fines of up to €50 million for systemic failure to delete illegal content. Supporters see the legislation as a necessary and efficient response to the threat of online hatred and extremism. Critics view it as an attempt to privatise a new 'draconian' censorship regime, forcing social media platforms to respond to this new painful liability with unnecessary takedowns. This study shows that the reality is in between these extremes. NetzDG has not provoked mass requests for takedowns. Nor has it forced internet platforms to adopt a 'take down, ask later' approach. At the same time, it remains uncertain whether NetzDG has achieved significant results in reaching its stated goal of preventing hate speech. This paper begins by explaining the background that led to the development and passage of NetzDG. It examines the reaction to the



law by civil society, platforms and the government. It concludes with suggestions, for platforms, civil society and the authorities, on ways to improve the law to be effective in the fight against online hate while keeping the internet open and free. CEPS acknowledges the Counter Extremism Project's support for this research. The study was conducted in complete independence. It is based on interviews with regulators, company representatives, and civil society activists. The authors take full responsibility for its findings.", "event-place": "Rochester, NY", "genre": "SSRN Scholarly Paper", "language": "en", "number": "ID 3300636", "publisher": "Social Science Research Network", "publisher-place": "Rochester, NY", "source": "papers.ssrn.com", "title": "Germany's NetzDG: A Key Test for Combatting Online Hate", "title-short": "Germany's NetzDG", "URL": "https://papers.ssrn.com/abstract=3300636"; "author": [{"family": "Echikson", "given": "William"}, {"family": "Knodt", "given": "Olivia"}], "accessed": {"date-parts": [{"2019", "12", "17"}]}, "issued": {"date-parts": [{"2018", "11", "22"}]}}, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}  
42 "Germany Lays down Marker for Online Hate Speech Laws," POLITICO, October 30, 2019.  
43 Echikson and Knodt, "Germany's NetzDG," or NetzDG law represents a key test for combatting hate speech on the internet. Under the law, which came into effect on January 1, 2018, online platforms face fines of up to €50 million for systemic failure to delete illegal content. Supporters see the legislation as a necessary and efficient response to the threat of online hatred and extremism. Critics view it as an attempt to privatise a new 'draconian' censorship regime, forcing social media platforms to respond to this new painful liability with unnecessary takedowns. This study shows that the reality is in between these extremes. NetzDG has not provoked mass requests for takedowns. Nor has it forced internet platforms to adopt a 'take down, ask later' approach. At the same time, it remains uncertain whether NetzDG has achieved significant results in reaching its stated goal of preventing hate speech. This paper begins by explaining the background that led to the development and passage of NetzDG. It examines the reaction to the law by civil society, platforms and the government. It concludes with suggestions, for platforms, civil society and the authorities, on ways to improve the law to be effective in the fight against online hate while keeping the internet open and free. CEPS acknowledges the Counter Extremism Project's support for this research. The study was conducted in complete independence. It is based on interviews with regulators, company representatives, and civil society activists. The authors take full responsibility for its findings.", "event-place": "Rochester, NY", "genre": "SSRN

Scholarly Paper", "language": "en", "number": "ID 3300636", "publisher": "Social Science Research Network", "publisher-place": "Rochester, NY", "source": "papers.ssrn.com", "title": "Germany's NetzDG: A Key Test for Combatting Online Hate", "title-short": "Germany's NetzDG", "URL": "https://papers.ssrn.com/abstract=3300636"; "author": [{"family": "Echikson", "given": "William"}, {"family": "Knodt", "given": "Olivia"}], "accessed": {"date-parts": [{"2019", "12", "17"}]}, "issued": {"date-parts": [{"2018", "11", "22"}]}}, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}  
44 "Final Report of the High Level Expert Group on Fake News and Online Disinformation," Text, Digital Single Market - European Commission, March 12, 2018.  
45 "Germany Proposes Europe's First Diversity Rules for Social Media Platforms," *Media@LSE* (blog), May 29, 2019.  
46 "Germany Proposes Europe's First Diversity Rules for Social Media Platforms."  
47 "France: Law on Manipulation of Information, Validated by the Constitutional Council, Is Published," accessed December 18, 2019.  
48 "Tackling Disinformation."  
49 "French Lawmakers Vote to Target Online Hate Speech in Draft Bill," Reuters, July 5, 2019. "container-title": "Reuters", "language": "en", "source": "www.reuters.com", "title": "French lawmakers vote to target online hate speech in draft bill", "URL": "https://www.reuters.com/article/us-france-tech-regulation-idUSKCN1U01UQ", "accessed": {"date-parts": [{"2019", "12", "18"}]}, "issued": {"date-parts": [{"2019", "7", "5"}]}}, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json"}  
50 "France's Law on Hate Speech Gets a Thumbs down," *EDRI* (blog), December 4, 2019.  
51 Kaye, "Letter of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression to the German Government," June 1, 2017.  
52 Article 595 and 278 of the Italian Criminal Code  
53 "Governments Countering Disinformation: The Case of Italy," *Disinfo Portal* (blog), November 20, 2019.  
54 Supra 40.  
55 Supra 40.  
56 Google Spain SL and Google Inc. v. Agencia Espanola de Protecci on de Datos (AEDP) and Mario Costeja Gonzalez, C 131/12  
57 Stichting Brein v Ziggo BV and XS4All Internet BV, C-610/15  
58 European Commission, recommendation on measures to effectively tackle illegal content online (last updated: 5 March 2018).  
59 "The CJEU's New Filtering Case, the Terrorist Content Regulation, and the Future of Filtering Mandates in the EU," accessed December 18, 2019.



- 60 ["The CJEU's New Filtering Case, the Terrorist Content Regulation, and the Future of Filtering Mandates in the EU,"](#) accessed December 18, 2019.
- 61 "Proposal for a Regulation of the European Parliament and of the Council on Preventing the Dissemination of Terrorist Content Online," Pub. L. No. 2018/0329(COD), 640 final COM(2018) (2018).
- 62 George Sadek, ["Egypt: Parliament Passes Amendments to Media and Press Law | Global Legal Monitor,"](#) web page, August 6, 2018.
- 63 Peter Guest, ["Singapore Says It's Fighting 'Fake News.' Journalists See a Ruse,"](#) The Atlantic, July 19, 2019.
- 64 ["Protection from Online Falsehoods and Manipulation Act 2019 - Singapore Statutes Online,"](#) accessed February 22, 2020.
- 65 ["Sri Lanka Proposes New Law on Fake News after Easter Attacks,"](#) France 24, June 5, 2019.
- 66 ["Fundamental Rights as Bycatch – Russia's Anti-Fake News Legislation,"](#) *Verfassungsblog* (blog), accessed December 18, 2019.
- 67 Marina Svensson, "Human Rights and the Internet in China: New Frontiers and Challenges," in *Handbook on Human Rights in China*, by Sarah Biddulph and Joshua Rosenzweig (Edward Elgar Publishing, 2019), 637, <https://doi.org/10.4337/9781786433688.00042>.
- 68 Operators of critical information infrastructure are an additional separate category of subjects, dealing largely with state activities.
- 69 ["China's Cybersecurity Law: What You Need to Know,"](#) accessed February 22, 2020.
- 70 "Translation."
- 71 ["China's New Internet-Censorship Rules Highlight Role of Algorithms,"](#) Wall Street Journal, accessed February 22, 2020.
- 72 ["China's New Internet-Censorship Rules Highlight Role of Algorithms,"](#) Wall Street Journal, accessed February 22, 2020.
- 73 Disinformation is defined by the High Level Expert Group report as "...including all forms of false, inaccurate, or misleading information designed, presented, and promoted to intentionally cause public harm or profit."
- 74 ["Introduction: Information Platforms and the Law,"](#) Georgetown Law Technology Review, July 21, 2018.
- 75 ["Regulating Disinformation with Artificial Intelligence,"](#) Study, Panel for the Future of Science and Technology, European Science-Media Hub: European Parliamentary Research Service, March 2019.
- 76 "Towards an internet Ombudsman institution" Provisional Report of the Committee on Culture, Science, Education and Media, Council of Europe, 9 December 2019





Prepared and published by the  
**NATO STRATEGIC COMMUNICATIONS**  
**CENTRE OF EXCELLENCE**

The NATO Strategic Communications Centre of Excellence (NATO StratCom COE) is a NATO accredited multi-national organisation that conducts research, publishes studies, and provides strategic communications training for government and military personnel. Our mission is to make a positive contribution to Alliance's understanding of strategic communications and to facilitate accurate, appropriate, and timely communication among its members as objectives and roles emerge and evolve in the rapidly changing information environment.

Operating since 2014, we have carried out significant research enhancing NATO nations' situational awareness of the information environment and have contributed to exercises and trainings with subject matter expertise.

[www.stratcomcoe.org](http://www.stratcomcoe.org) | [@stratcomcoe](https://twitter.com/stratcomcoe) | [info@stratcomcoe.org](mailto:info@stratcomcoe.org)